DOCUMENT RESUME

ED 298 968 IR 052 509

AUTHOR Kahn, Paul
TITLE Information Retrieval as Hypermedia: An Outline of InterBrowse.
PUB DATE [88]
NOTE 10p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer System Design; *Database Management Systems; *Databases; *Information Retrieval; *Man Machine Systems; Microcomputers; Online Searching; *Search Strategies
IDENTIFIERS Gateway Systems; *Hypermedia

ABSTRACT

InterBrowse, a uniform interface information retrieval application for several different databases, is designed to be used in Intermedia, a hypermedia environment currently under development at Brown University's Institute for Research Information and Scholarship. This application arose out of the recognized need for an interface that can be used with the hundreds of collections of bibliographic and numeric data, reference collections, and collections of digital images that are currently available on the software market. The provision of access to this electronic information is viewed as a continuum of need and use. Locating and applying information, i.e., integrating information, must be viewed as a continuum of interlocking tasks, beginning with a perceived information need and resulting with that needed information being applied to the user's work environment. The application used here is broken down into five steps: (1) a high-level browser; (2) a database-specific browser and query builder; (3) a set of tools for refining and reapplying queries; (4) a tool for using results to develop queries in other collections; and (5) a personal management information system. The linking structure of the hypermedia environment used here adds another layer of possible information management: saving links among pieces of information in a way that makes these links easy to navigate. Two figures are included. (5 bibliographic notes/references) (Author/EW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

INFORMATION RETRIEVAL AS HYPERMEDIA: AN OUTLINE OF INTERBROWSE

Paul Kahn, Brown University

ED298968

The need to provide access to electronic information from within the daily environment of a knowledge worker presents many interesting challenges. The current software market contains hundreds of collections of bibliographic and numeric data, reference collections, and collections of digital images, all in discrete and proprietary formats. At the same time, many database management systems are being employed to reorganize and make use of data collected from these discrete sources. A uniform interface to these database resources is needed. To test out our ideas in this area, we are developing an information retrieval application that will run within Intermedia, the hypermedia environment currently under development at Brown's Institute for Research in Information and Scholarship (IRIS). This application, called InterBrowse, will address the issues of information retrieval and management in a hypermedia environment. This paper provides an outline of the requirements for that application

We view this problem as a continuum of need and use. Locating and applying information, i.e. *integrating information*, must be viewed as a continuum of interlocking tasks, beginning with a perceived information need and resulting in that needed information being applied in the user's work environment. We break the solution down into five steps: 1) a high-level browser, 2) a database specific browser and query builder, 3) a set of tools for refining and reapplying queries, 4) a tool for using results to develop queries in other collections, 5) and a personal information management system. The linking structure of our hypermedia environment adds another layer of possible information management: saving links among pieces of information in a way that makes these links easy to navigate and apply.

1 ASSUMPTIONS

Personal computers are being used today to create *personal databases* in which information is gathered, stored, referenced, and manipulated by a single user. When access to large databases via timesharing networks began in the late 1970s, the end-user worked at a terminal, either block mode or teletype in style, connected to a host computer over a telecommunications network. Today, terminal emulation and special purpose front-end software on personal computers is commonplace for searching these same on-line databases. These applications, however, solve only a small part of the problems faced by end users of large databases. They automate a number of repetitive tasks, such as interacting with a telecommunications network (Telenet, Tymnet, etc.), logging on to a timesharing system, and even uploading pre-created queries. But they do not deal well with issues such as: presenting tools to help a user build a reasonable query, resolving queries among different database collections, providing graphical representations of result sets, or downloading information into a personal database.

IR052509

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Paul Kahn
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

As 1988 begins the personal computer field is advancing rapidly in three areas:

CPU power and memory: We now have a generation of machines based on a 32-bit CPU architecture (80386, 68020 and 68030, RISC), capable of addressing tens of megabytes RAM and usually configured with 4-10 megabytes of RAM.

Size of data storage: The standard magnetic diskette is now 1-2 megabytes, with most local hard disks in the 80-140 megabyte range. Optical disk options include the 550-600 megabyte CD ROM and WORM cartridges in the 200-800 megabyte ranges.

Size and resolution of display: Display size is increasing on all fronts, with 15-19" monitors, minimum resolutions of 75 dots-per-inch, and 256-1024 colors.

Telecommunications: The 300/1200 baud standards are giving way to higher speeds such as 9600-19,200 baud modems.

The current applications software market lags far behind the potential offered by these advances in computing and storage hardware. The large capacity of an inexpensive optical media such as CD ROM provides the opportunity to deliver very large and complex databases to single users. To take advantage of this new media, however, equally large advances must be made in user interface design.

Since 1983 with the introduction of the Apple Lisa followed by the original 128K Macintosh, windowing systems and pointing devices have clearly demonstrated their ability to increase the potential user base for all applications.

Our experiences with numerous hypermedia projects at Brown have also demonstrated that the use of graphics makes it easier for users to visualize, analyze and manipulate complex information.

2 THE CONTINUUM MODEL

Integrating information must be viewed as a continuum of interlocking tasks, beginning with a perceived information need and resulting in the needed information being placed in the user's work environment. To do this an application must be capable of presenting information about heterogeneous database resources in some open-ended manner. It must be able to deal with the user's needs from beginning to end, an end that is not reached by simply displaying information on a screen. The more complex issues of storing and organizing results in a personal information management must be part of the information retrieval design.

To test our ideas in this area, we are developing an information retrieval application that will run within Intermedia, the hypermedia environment currently under development at Brown's Institute for Research in Information and Scholarship (IRIS) [1]. This application, called InterBrowse, will address the issues of information retrieval and management in a hypermedia environment. For the design of InterBrowse, we are breaking the application into five tasks which are capable of sharing and transferring information in a simple and fluid manner. We identify these as serving the following functions:

1) a high-level browser for viewing collections of available information leading to the selection of a database;

2) a database specific browser and query builder for browsing information in a specific database and developing a query;

3) a set of tools for refining and reapplying queries, for browsing the results of a query to refine and reapply those results to another query, to narrow, broaden, or change the direction of a search, and select from those results the desired information;

4) a tool for using selected results to develop further queries in other database collections available through the high-level browser;

5) a personal information management system to store results locally in a systematic fashion for use in the work environment.

Figure 1 provides a conceptual overview of these interlocking tools. Each of these are described in more detail below.

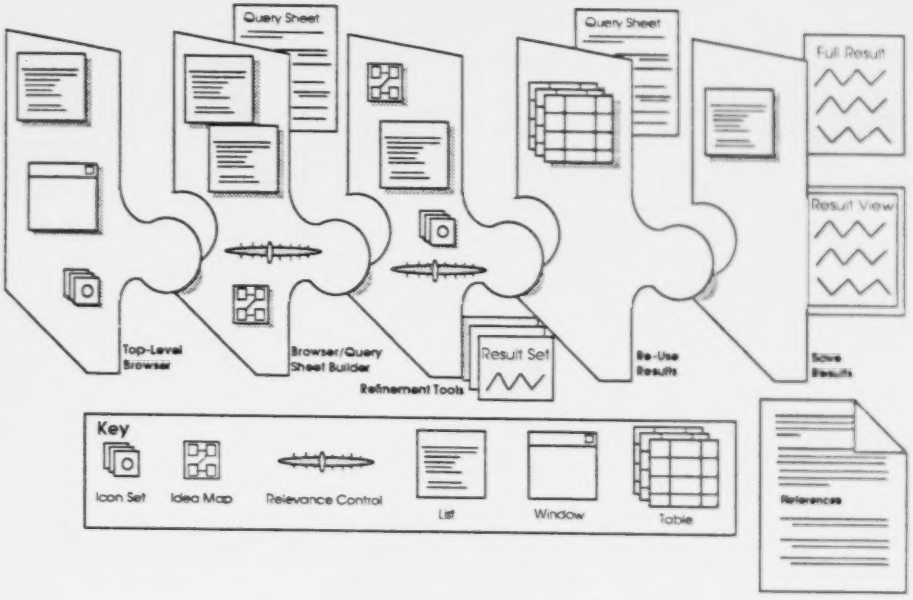

Figure 1: An overview of InterBrowse

2.1 High-Level Browser: Browsing collections of available information leading to the selection of a database

The purpose of this browser is to provide the user with information about the available databases. The browser will present this information in a number of ways, analogous to the way the Macintosh Finder presents file/folder information, i.e. as groups of icons or as sorted lists of names.

An interesting example of this kind of database browser is to be found in Pro-Search, a personal computer application for searching databases on BRS and Dialog.[2] This program provides a high-level interface to the several hundred databases found on these information services. The user walks a tree by selecting one of four general topics (*Art and Social Science*), one of a list of subtopics (*Language*), and is finally presented with a set of "cards" representing the databases that fall under these categories (*Arts & Humanities Search, Language & Language Behavior Abs, MLA Bibliography, etc.*). The same database could be reached from a different subtopic (*Philosophy* also leads to *Arts & Humanities Search*). InterBrowse will provide this functionality using a methodology similar to the "View" function in the Macintosh Finder.

Selecting a database name or icon will allow the user to get information about that database. This will operate in a manner analogous to the "Get Info" function of the Macintosh Finder. Information accessible from any database name or icon will include:

- size, coverage and time-depth of the database, in terms of general subject covered, list of information sources, frequency of update, etc.

- overview of the field structure of the database, showing how records are organized, what is searchable, etc.

- description of the thesaurus, the citation index, or other special features, explaining how subject headings are organized in different structures: flat lists, hierarchical trees, or networks of greater and lesser terms

- a user's or group's history of interaction with this database

The user can then choose to develop a query for one database by following from the information window to the query sheet builder.

2.2 Query Sheet Builder: Browsing information in a specific database leading to the development of a query

The purpose of this tool is to provide the end-user with the means for examining the structure and content of a particular database, and applying that knowledge to developing a query in that collection. The tool provides information about:

- what the database contains

- how the database defines a particular subject

- how the subjects relate to the searcher's ideas

By interacting with the Query Sheet Builder, the user creates a document called a Query Sheet, which can then be used to generate a result set from a database. To do this InterBrowse must know how each database is structured and indexed, and how to request results from the appropriate information retrieval (IR) system.

The process of building a Query Sheet involves two paradigms: *selection* from a sorted list or graphic and *feedback* resulting from that selection in the form of numbers, re-sorted lists of result terms, or graphics.[3]

The Query Sheet Builder will provide graphic *idea maps* showing coverage of subjects, relations among subjects, bibliometric features of the data, and co-citation cluster maps showing central works. Things we would like to model include:

- a view of an entire database or a large subset as a topological *idea map* which presents the coverage of subject areas as surface plots. The tools for calculating the "shape" of these different visualizations will allow the user to select specific areas for further searching.

- the use of co-citation information to calculate directed graphs for locating the "central" authors, subject headings, journals, and citations for a specific search, subject, or combination of subjects.

- graphic representations of bibliometric functions such as author and journal popularity for the entire database.

To find out what the database contains the user will have access to sorted lists of the contents for selected fields such as authors, subject headings, and journal titles. This feature will allow the user to identify, for example, whether the work of a particular author is contained in the database by examining the author list for the author's name. If it is there, the user can then select the correct form of the author's name to build the query. This technique allows the user to build a query by selecting items from the field indices, rather than having to guess the form in which a term exists or whether it exists at all. If the user chooses to supply terms directly from the keyboard, the Query Sheet Builder will parse the input, and match words against indexed terms. There will be a mechanism for notifying the user when input does not match any indexed terms, and this mechanism will allow the user to browse the appropriate list for alternatives.

The way in which a database defines its subjects is best presented in the form of a subject thesaurus. An interesting example of thesaurus access for a single database is found in the Knowledge Finder version of MEDLINE, a Macintosh application which accesses MEDLINE on CD ROM.[4] This application treats the Medical Subject Heading thesaurus (MeSH) as a database to aid the user in locating subject headings appropriate to a query. Usage and history notes for each term, supplied by the National Library of Medicine, are available for display in separate windows. InterBrowse will treat the thesaurus of each database in a similar fashion. Relations among subject headings will be represented in a manner which allows the user to navigate among the connections, so that a user can move among and select networks of related terms.

By supporting multiple windows on a large (19") monitor, InterBrowse will allow the user to maintain visual context among lists of subject headings, information about these headings, list of terms in various fields, while still viewing the current Query Sheet. As terms are selected from these lists a query is built. Once the user is satisfied with the initial query the Query Sheet is *applied* to the database.

2.3 Query Refiner: Using the feedback from a query to refine and reapply, and/or select from those results

The result of applying the Query Sheet will be the creation of a result set. The purpose of the Query Refiner is to provide tools for evaluating the result set and using the information it contains to refine and reapply the Query Sheet.

It is assumed that coarse results (large sets) are just as useful as fine results (small sets) when the purpose is to refine and reapply a query. The Query Refiner will provide tools to browse and represent large or small result sets.

A fundamental aspect of the InterBrowse design which distinguishes it from previous applications is its ability to deal with result sets as sub-databases. For the sake of efficiency, most IR systems produce a result set which consists of a set of pointers to records which satisfy a query. To effectively represent a result set for browsing, InterBrowse requests the IR system to return *all* the information in, not just pointers to, the records which satisfy the query. The Query Refiner tool will receive the entire result set and organize the information as a sub-database for display and refinement purposes.

The Query Refiner will present the following information about the result set:

- sorted lists of the result set's contents for selected fields such as authors, subject headings, journal titles.
- graphic *idea maps* analogous to the pre-calculated maps of the entire database available in the Query Sheet Builder.
- search results as a group of *graphic icons* which visualize features that the user deems important. The graphic attributes of the icon set can be based on applied bibliometrics such as author and journal frequency, numbers of subject headings (generalness or narrowness), etc.[5] The relative weights and graphic presentation of these attributes will be controlled by the user.
- when citation information is included in the result set, the Query Refiner will generate co-citation cluster maps to show central works in a result set.

Each of these displays will provide an easy way for the user to select information to modify, add to, or otherwise refine the current Query Sheet, and to reapply it to create a new result set.

During a session, InterBrowse will keep a version history of the Query Sheet so that the user is free to make and "undo" modifications to the search. Any version of a Query Sheet can be saved as an InterBrowse document.

Two useful *filters* will be provided as options for limiting the view of any result set: Local Holdings and Full-text Online.

The Local Holdings filter will limit the display to only those sources found in the user's library. For journals this can be done by filtering against a list of the serial holdings of the local library. For books it can be done through access to the data in an online union catalog, such as Brown's online catalog, JOSIAH, or a multi-site union catalog such as RLIN. This will help the user evaluate results based on sources that can actually be located.

The Full-text Online will limit a display to those citations which can be used as pointers to the full-text of an article or book in electronic form. This will bring to the user's attention those citations that can be located online for further searching and browsing.

It will be possible, while browsing any of these displays, to select individual records or an entire result set to be used in the Correspondence Builder, and/or saved in the Personal Information System, both of which are described below.

2.4 Correspondence Builder: Make use of results to develop a query in another collection

The InterBrowse design assumes that the information a user needs may be located in any of a number of databases, and that each database may represent similar ideas (subject headings) and identical elements (titles, author names, journal names) in dissimilar ways. The purpose of the Correspondence Builder is to help the user locate information that overlaps with, or is related to, information in saved result sets. It will provide another set of tools that can be applied to the Query Sheet and result sets:

- tools for reapplying a Query Sheet to another database
- tools for showing relations among results from various databases

Tools will be provided for building tables to show the relations between database-supplied subject headings in selected results and user-supplied subject headings assigned to those records in the Personal Information Manager.

For example, a user might be looking for articles about how database software is used to manage bibliographic information. Previous results from a search for information on the concept of *bibliographic data bases* in both ERIC and the MLA Index show that this phrase exists as a subject heading in both databases. In ERIC most of the thirty-three records found also have subject headings such as *online searching* and *databases*. In the MLA the two records with the subject heading *bibliographic data bases* also use the subject heading *research tools*. By browsing the MLA, the user locates three more interesting records with *research tools* and *data base*, and nine others with *data base* and *computer-assisted research* as subject headings. As the user copies each of these records into his Personal Information Manager, the Correspondence Builder will map the user-assigned keywords to the database-supplied subject headings. Based on the results of previous searches, this tool will be able to suggest possible synonyms and related subject headings in these two databases for later searches.

This tool will maintain a table of what sources are indexed in what databases. This will allow the user to look for information that is likely to be found in specific journals. It will also provide the ability to use bibliographic citations from one database as "pointers" to full text versions of an article available in another collection, as described in the Full-text Online filter above.

2.5 Personal Information Manager: Storing results locally in a systematic fashion for use in the work environment

We assume that the user will want to save and use information located in a result set. The Personal Information Manager will be capable of accepting selected "pieces" of information from a result set, and will provide tools to help the user organize the copy of this information.

A result set will be treated as a group of *objects* which the user can directly manipulate. Selecting a result will involve marking or moving a result object into a *save* space.

Transferring (copying) information to the local system will be done in a form that retains the semantic

information in each record, and either transfers the complete record or enough information to retrieve more information when necessary.

In citation collections, InterBrowse will bring back full tagged records complete with database-specific subject headings and identifiers for *classification*, along with a selected view of that record for *display*. In full-text, InterBrowse will bring back sentences or paragraphs, as defined by the user, and enough information to return to the complete article if necessary. In numeric collections, InterBrowse will bring back data along with semantic relations to other data in the collection.[6]

Saved results will become part of a local database, built by the Personal Information Manager, which will contain tools for normalizing data in the personal or group work space. These tools will include:

- correlation tables for forms of author names and journal titles
- mapping personal subject classifications to database-specific subject headings

The Personal Information Manager will be capable of integrating information into the text/graphics authoring work environment of other Intermedia applications. It can be used to generate sets of footnotes, bibliographies, or other forms of citation for use with other Intermedia editors.

3 INFORMATION RETRIEVAL AS HYPERMEDIA

Current IR systems do not provide the flexible linking structure characteristic of hypertext/hypermedia systems. Having located an instance of the string "Oliver North" in a newspaper database, there is no way for a user to create a link between that string and the string "William Casey" in another story. At the same time, the current family of hypertext/hypermedia systems do not provide the basic record/field/string retrieval functionality so basic to information retrieval.[7] InterBrowse will provide a bridge between these two domains, providing access to a variety of IR systems from within Intermedia, a sophisticated hypermedia framework.

Using Intermedia, readers can browse through linked information in a non-sequential but orderly manner. Information is entered into Intermedia using one of a variety of applications, which include *InterText*, a text processor; *InterDraw*, a graphics editor; *InterVal*, a timeline editor that for organizing temporal information; *InterPix*, a scanned-image viewer; and *InterSpect*, an application to view three-dimensional models. Figure 2 illustrates an Intermedia session.

Intermedia is both an author's tool and a reader's tool. The system, in fact, makes no distinction between types of users, provided they have appropriate access rights to the material they wish to edit, explore, or create. Creating new materials, such as an InterBrowse Query Sheet, and making and following links are all integrated into a single seamless multi-user environment.

To link information, the user indicates a source selection (a *block*) in a document made with any of these editors and chooses the Start Link command. After performing any number of related or unrelated actions, the user indicates a destination selection (another *block*) in the same or some other document, and chooses the Complete Relation command. The system places a link marker beside each selection. When the operation is finished there is a semantic tie — a navigational link — between the two selections. Whenever the user selects a link marker and issues the Follow command, the document containing the other end of the link is activated and the appropriate information is presented in another window.

Intermedia stores link information is stored in a *web* database, not in the file itself. One implication of this strategy is that any user can create links to otherwise read-only documents, such as an online dictionary or citation database, without in any way corrupting the data itself. Webs present users with a context in which to collect and navigate a set of links. An individual user, a group of users, or even an entire campus can conceivably create a shared web of linked information.

Because InterBrowse is incorporated into Intermedia, users are able to link to and from search results stored as InterBrowse documents. These search results will be stored in two forms:

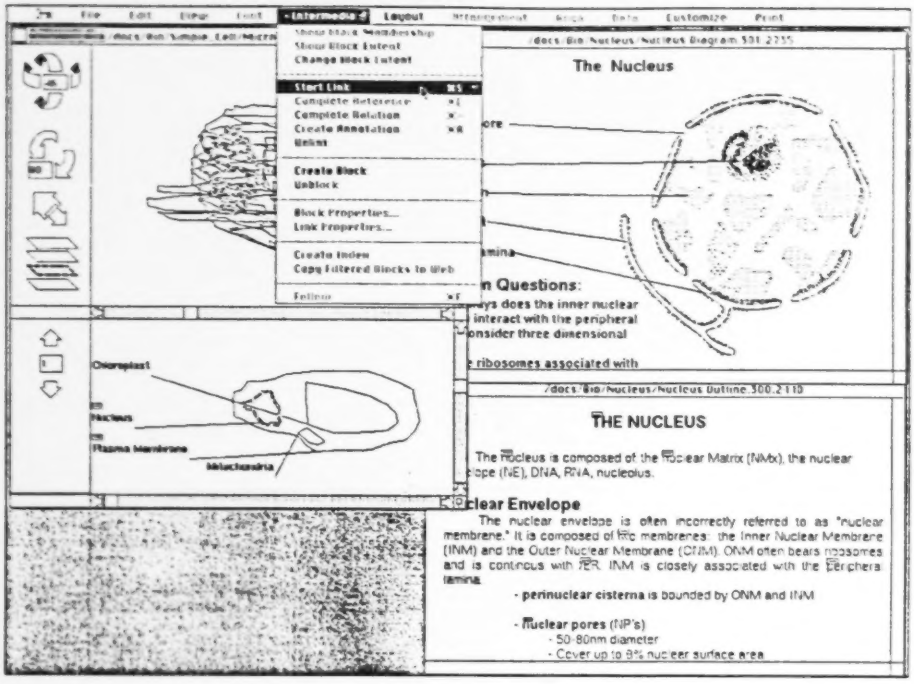
Start Link
Create Block
Unblock
Block Properties...
Link Properties...
Create Index
Copy Filtered Blocks To Web
The Nucleus
Questions:
ribosomes associated with
Chloroplast
Nucleus
Plasma Membrane
Mitochondria
THE NUCLEUS
The nucleus is composed of the nuclear Matrix (NMx), the nuclear
lope (NE), DNA, RNA, nucleolus.
clear Envelope
The nuclear envelope is often incorrectly referred to as "nuclear
membrane." It is composed of two membranes: the Inner Nuclear Membrane
(INM) and the Outer Nuclear Membrane (ONM). ONM often bears ribosomes
and is continous with RER. INM is closely associated with the peripheral
lamina.
- perinuclear cisterna is bounded by ONM and INM
- nuclear pores (NP's)
- 50-80nm diameter
- Cover up to 8% nuclear surface area

Figure 2: An Intermedia session. Links are created between the two-dimensional bitmap of a cell nucleus on the upper right, the three-dimensional wireframe of the cell on the left, and the text document on the lower right.

- as a Query Sheet containing the algorithm for locating information: the search path to database, and the actual search query
- as a Result sets stored in the Personal Information Manager

The Intermedia linking mechanism is used in InterBrowse to save or create links among pieces of information to make it easy for the user to follow and navigate among related items. InterBrowse will allow for two kinds of links: block to block link and action link. An overview of the linking strucutre is found in Figure 3. Of these, the first is currently implemented in all Intermedia applications. The other is under development.

3.1 Block to Block Link

A block to block link is a bidirectional link between two blocks in any Intermedia document. These links are created manually by the user selecting two blocks and making the link. InterBrowse will use block to block links for two purposes:

Link between a block in any document and a Query Sheet for *viewing* or *editing*:

Query Sheets can be saved and linked to from any Intermedia document for later reference and modification. Following a link from a text or graphics document to a Query Sheet will start InterBrowse and display the Query Sheet. Similarly, a block in a Query Sheet can be linked to another Query Sheet, or any other Intermedia document.

Link between a block in any document and a view of one or more references stored in the Personal Information Manager:

A link can be made from a any document to a view of a citation. Following this link will open the view of the citation in the Personal Information Manager. Following the link from the citation will bring up another document, or citation.

3.2 Action Link

An action link is a uni-directional link starting from a block and causing some action to take place. The action is specified by information in the document which is the target of the link. InterBrowse supports action links by linking a block in any Intermedia document to a Query Sheet, which is applied with the link is followed. InterBrowse will use action links for two purposes:

Link from a document to a Query Sheet

A link can be made from any block to an existing Query Sheet. Following this link will apply the query, produce a result set and present it to the user in the InterBrowse Query Refiner

Link from the Personal Information Manager to a Query Sheet

A link can be made from a reference stored in the Personal Information Manager to the database from which it came. Following this link goes back to the stored Query Sheet and re-locates the reference in the source database with the InterBrowse Query Refiner. The Refiner tools can then be used to continue browsing this database.

4 CONCLUSION

The design of InterBrowse addresses a large number of issues facing both information retrieval and hypermedia research. The design is based on our speculation, not that these issues need to be addressed, but that they can now be addressed by the methods we have used to create other scholarly tools in Intermedia. Given the advances in personal computer hardware, user interface design, and programming techniques, we feel that the time is right.

The development of InterBrowse will be accomplished in stages over the next few years. It is our hope that this research will prove valuable for other researchers in the field, as well as members of the the Online community at large.

Submitted 21 January 1988. The author wishes to acknowledge the contributions of other members of the InterBrowse design team: Marty Michel, James H. Coombs, and Karen E. Smith.

Notes:

1. The design and functionality of Intermedia are described in a number of articles, the most recent of which is "Intermedia: The Concept and the Construction of a Seamless Information Environment", N. Yankelovich, et al., *IEEE Computer*, January, 1988, pp. 81-96.
2. *Pro-Search* manual, Ann Arbor: Personal Bibliographic Software, 1986.
3. In "Making A Difference: A Review of the User Interface Features in Six CD ROM Database Products" (*Optical InformationSystems Journal* 8:3 May/June 1988) I discuss the selection and feedback aspects of some current CD ROM database applications.
4. *Knowledge Finder* manual, North Andover: Aries Systems Corporation, 1987.
5. The Graphical Icon-driven Search Tool (GIST), a prototype for this kind of graphical representation of bibliographic data, is described in "Graphic Icons for Problem-Oriented Information Retrieval", Page Elmore, Providence: Brown University, IRIS Technical Report 87-1.
6. The initial version of InterBrowse will be limited to handling only bibliographic citation data.
7. See "Hypertext: An Introduction and Survey" J. Conklin, *IEEE Computer*, Sept. 1987 for a useful survey of such systems.